Probabilistic Models for Computer Architectures

Adolf Filin, Vadim A. Malyshev, Anatolij Manita

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Probabilistic Models for Computer Architectures

Adolf FILIN - Vadim MALYSHEV
Anatolij MANITA

N° 2419
Novembre 1994

PROGRAMME 1

Rapport de recherche

Modèles Probabilistes pour les Architectures des Ordinateurs

Adolf Filin* Vadim Malyshev† Anatolij Manita‡

Novembre 1994

Les relations entre les files d'attente avec priorités et les architectures des ordinateurs sont bien connues. Mais, à notre connaissance, il n'y a pas de modèles exacts englobant toutes les architectures. Cette article a deux buts: le premier est de donner une formulation en termes mathématiques; le deuxiéme (et le plus important) est de présenter une nouvelle approche pour traiter les systèmes de files d'attente avec priorités.

*Institut de Problemes d'Informatique, Moscow, Russie

†I.N.R.I.A.—Domaine de Voluceau—Rocquencourt—B.P.105-78153, Le Chesnay Cedex, France

‡Laboratoire de Grands Systèmes Aléatoires, Université de Moscou, Moscou, 119899, Russie

Probabilistic Models for Computer Architectures

A.V.Filin*   V.A.Malyshev†   A.D.Manita‡

November 1994

Abstract

Connections between priority queueing models and computer architectures are widely known. But, as far as we know, there was no formulations of exact models for sufficiently general computer architecture models. This paper has two goal: the first and the smaller one is just to give this formulation in exact mathematical terms. The second and the most important one is to present a new approach to priority networks themselves. This approach is based on recent advances in the dynamical system approach to queueing networks, which in some very particular cases becomes some well-known fluid approximation. This gives a new approach to evaluate performance of a given computer architecture. We apply this method here to the simplest architecture with the unique bus. This paper can be considered as the first step in the development of this approach. Next steps are: more complicated architectures, optimization (one optimization problem is considered here) etc.

*Postal address: Institute of Problems of Informatics, Moscow, Russia

†Postal address: I.N.R.I.A.—Domaine de Voluceau—Rocquencourt—B.P.105–78153, Le Chesnay Cedex, France

‡Postal address: Laboratory of Large Random Systems, Moscow State University, Moscow, 119899, Russia

1 Introduction

The goal of this paper and of the following ones is to formulate some global probabilistic models for computer architectures. Mathematical modeling of computer architectures presupposes knowledge of statistical structure of all data flows between different parts of the computer. Everyone understands that the exact knowledge of such statistics is impossible, because it strongly depends on the problem that the computer solves at the moment and moreover it depends also on the "psychology" of the person who has written the program for solving given problem. But nevertheless some models seem to be useful if

i) they give rise to interesting mathematics;

ii) they give qualitative understanding of some practical phenomena, that could be of interest for computer designers;

iii) assuming that books on computer architecture are unreadable for mathematicians, global exact models formulated on a rigorous level, can make this field more attractive for mathematicians.

Computer can be a very large and complicated system and can be looked at from various points of view and also at different scales.

The scale we choose here is the most detailed one: hardware and microprogramming scale.

The point of view we choose here is to present the computer as a *large system of priority queues.* So our models are different from models overviewed in paper [1] which deals with nonprioritive disciplines: first-come-first-served, time sharing, no queueing, last-come-first-served. In our exposition we mix nonrigorous exposition and rigorous models. As our exposition is intended for mathematicians we explain some points which are trivial for specialists in computer architectures.

We present here some models in the order of increasing complexity. Sections 2.1 and 2.2 are based on hardware or microprogramming level. Section 2.3 contains already programming level. In the mathematical part of the paper we have managed (due to time constraints) to analyze only the simplest computer architecture (the unique bus, see Fig.1)

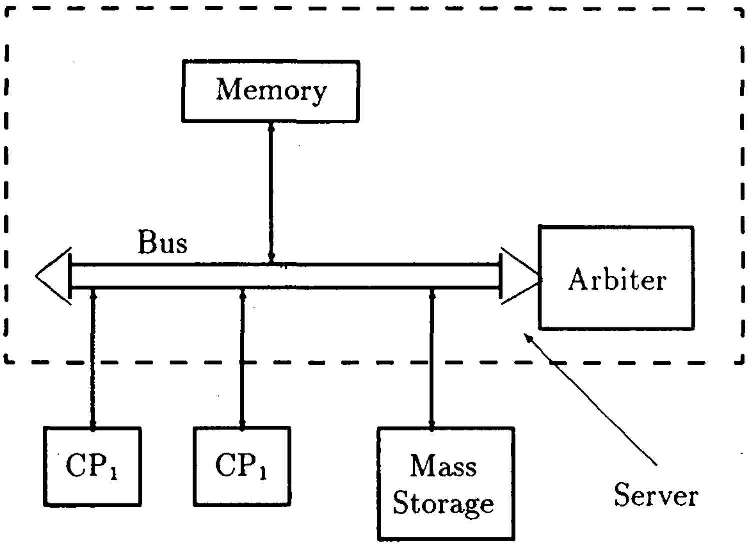

Figure 1: Structure of a computer with unique bus

2 Direct Memory Access Models

2.1 Simplest Model

The following Figure 1 illustrates the unique bus structure shared by all the units in the computer

The properties of this structure (Model 1) are the following:

• The memory here is the main memory. Communication can take place between memory and any other unit in both directions.

• Only one communication act (memory $\longleftrightarrow$ other unit) can take place at a time, because there is only one resource (bus) for this communication.

• Units A_i can be central or peripheral processors, mass storage, streamer, display processors, I/O devices (a device is called a direct access device if it can control by itself, without a processor, memory-memory exchange).

• The system "memory-bus-arbiter" is considered as a server in the sense of queueing theory and units A_i as the sources sending demands to this server. Our central assumption is that

Assumption 1 *All these sources send their demands independently of each other.*

Strictly speaking, this assumption is always wrong. For example, depending on the program which the processor executes, the memory should take something from mass storage. It contradicts to that we assume mass storage sending its demands independently. But it is clear that this approximation can be reasonable in some cases.

• There exists some priority discipline P_α. It can be for example relative priority P_{rel}, absolute priority P_{abs} (with preemptive resume or not) or others. Normally, in this model P_{rel} is taken for granted. To each unit A_i some priority number $P(i)$ is prescribed. The problem is to find the function $P(i)$ which fits better to some important parameters of the computer system.

• Some of the units A_i are considered to be "infinite sources", that is they send an infinite sequence of unit demands. Some of them are considered to be "finite sources" where each demand from node i tries to reach the server (at node 0) independently of the other such demands. There are finite number of such demands.

• The main parameters are following:

(a) for the system as a whole: the distributions of the busy and nonbusy periods and length of the queue;

(b) for all units also the waiting time of a demand.

Now we describe *the mathematical queueing model* (see Figure 2) in more exact terms.

We consider BCMP-type network of mixed (closed-open) structure (see [1]). There is one "central" node 0 and k other nodes $1, \ldots, k$ (which correspond to finite sources). For each $i = 1, \ldots, k$ there are N_i customers, m_i of which are situated in the queue to node 0 and the other $N - m_i$ at the node i. Each customer from node i, independently of the others goes to node 0 with exponential distribution of intensity λ_i and goes backwards to node i after being served at node 0. The discipline at node 0 is some P_α.

There are also incoming flows $k+1, \ldots, N$. They are assumed to be Poisson with intensities λ_i. There can be also cases (timers, ...) when demands come in some regular sequence. They are put to the queue of node 0 and are served in accordance with the discipline P_α. After being served they quit the system.

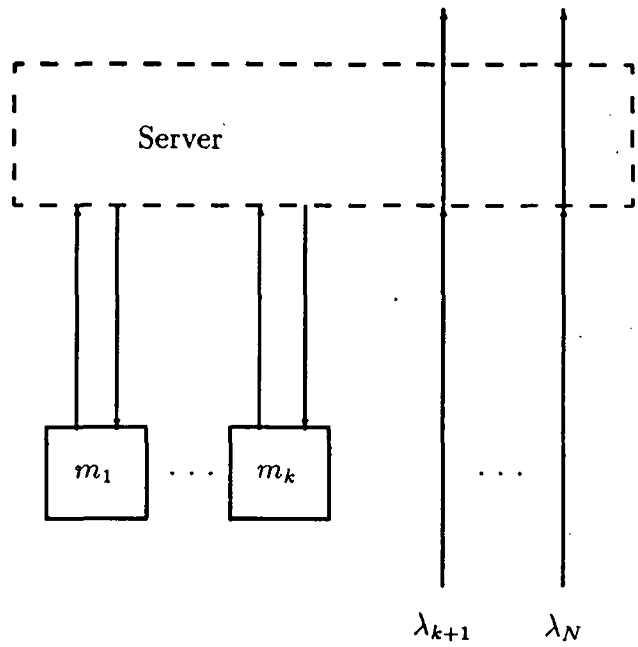

Figure 2: Mathematical queueing model

Service times at node 0 of the demands of type i are independent with an (arbitrary) distribution S_i.

Priority disciplines.

Any service discipline is a rule which determines which of customers waiting for service should occupy the server when the server node is ready to serve a customer. Priority disciplines constitute a special class of disciplines which assume that there are different customer types and these type are of different importance. Usually one enumerates customer types by natural numbers in the order of decreasing importance: type i has priority with respect to type j iff $i < j$. One distinguishes *absolute* and *relative* priority disciplines. Difference between these two prescriptions consists in what happens in situation when the server is occupied by some customer but at the same time moment another customer with higher priority arrives.

In case of *absolute* priority discipline customer with high priority interrupts service of customer with lower priority and immediately occupies server node.

1. Discipline is called absolute priority *with preemptive resume* if inter-

rupted customer should return to server node (when system becomes free of customers of more high priority) and be served during the time remained after the previous service of the customer.

2. Discipline is called absolute priority *without preemptive resume* if interrupted customer leaves the system.

In case of *relative* priority a customer with higher priority waits until server will finish to serve the customer with lower priority.

2.2 Matrix Model

Here the situation is the same as earlier but the units have double indices (see Figure 3)

$$A_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, N_i;\quad \sum_{i=1}^{n} N_i = N.$$

Priorities are in the lexicographic order. Index i defines the level, priorities between levels are assumed to be absolute with preemptive resume (the most often case) or sometimes relative. Inside the levels the priority discipline is relative priority.

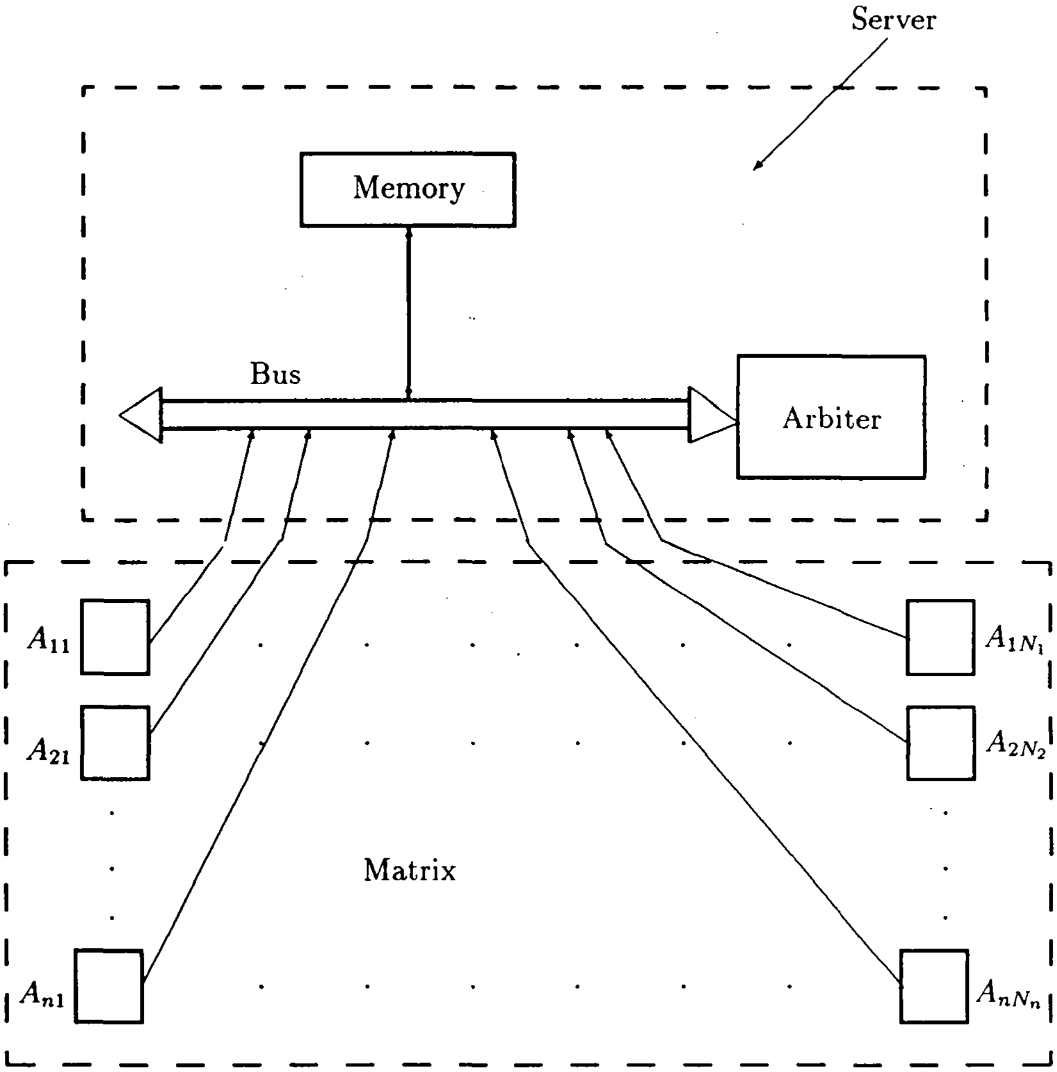

Figure 3: Matrix model

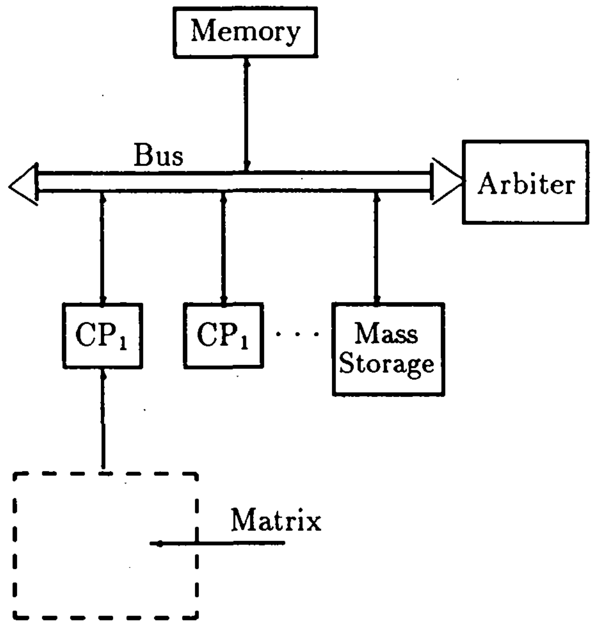

Figure 4: System with a subsystem of programming devices

2.3 Direct Access System with Programming Subsystems

Let us consider the following model of *direct access system with simplest subsystem of programming devices* (see Figure 4).

So we are in the situation of Figure 1 but, for one of the units (processor A_1), there is a matrix subsystem. This processor A_1 is a server for this matrix subsystem, which has the same priority rules as it was indicated in subsection 2.2. Simultaneously the processor A_1 sends demands to the bus (the main server).

Assumption 2 *The processor A_1 sends demands to the bus only while working as a server (i.e. when there is a queue to A_1). Otherwise its demands are independent of other details (which device a_{ij} is served by the processor, what kind of demands go to this processor from a_{ij} etc.). In other words, the processor becomes a (finite or infinite) source during the time it is being a server.*

This assumption is also wrong in all cases but can be a rough approximation and can give some qualitative information. If A_1 is a finite source then this assumption needs to be stated more exactly and this can be done in different ways.

2.4 Several Subsystems of Programming Devices

The Figure 5 shows more complicated system which is obtained just by concatenation of simplest ones from the previous subsection.

So, some of A_i can be servers for matrix subsystems $M(i) = (a^i_{jk})$. Some of a^i_{jk} (if it is a processor) can be servers for other matrix subsystems $M(i;j,k)$ and so on by induction. But it is assumed that any processor can be a server for at most one matrix subsystem. So we have situation of Figure 6.

So we have finite chains of processors $B_0,\ A_i = B_1,\ B_2, \ldots, B_n$, where B_0 is the main server (memory-bus-arbiter), and B_i is the server for $B_{i+1}, i = 0, \ldots, n-1$. The main assumption is similar to the Assumption 2.

Assumption 3 *The processor B_i can send demands to processor B_{i-1} only while working as a server (i.e. when there is a nonempty queue of demands from B_{i+1}).*

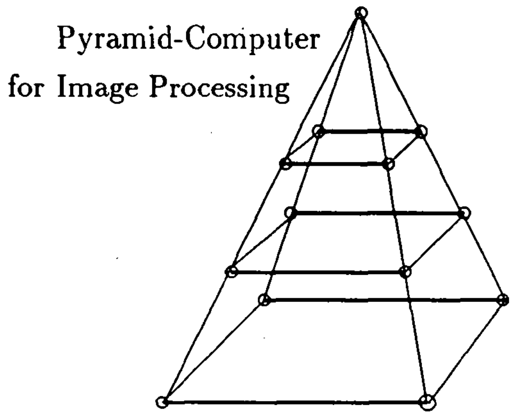

Figure 5: Pyramid

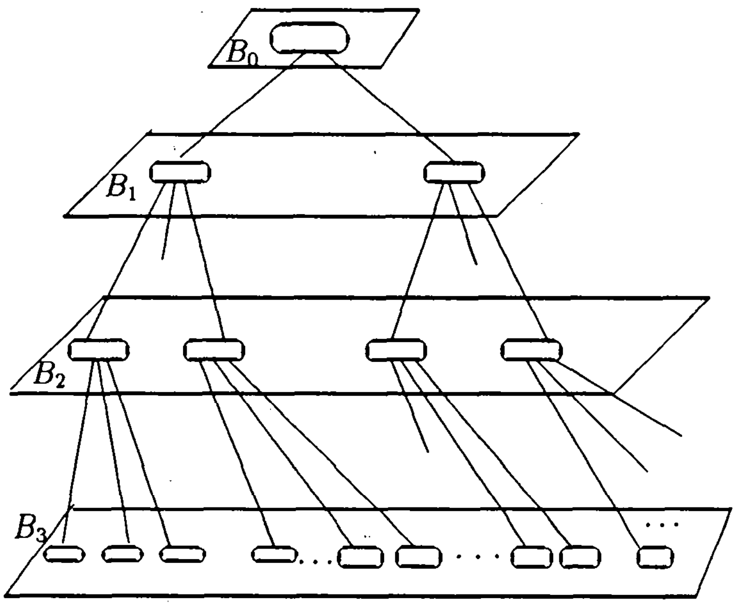

Figure 6: Hierarchy of matrix subsystems

3 Probabilistic Analysis

We shall deal here with model of section 2.1.

This section is organized as follows. In section 3.1 we give rough analysis of system of infinite sources with absolute priority with preemptive resume. We do not make any restrictive assumptions concerning input flows and service time.

In the next sections we make some assumptions about input flows and service time to be able to use Markovian description of priority systems. We consider the following disciplines: absolute priority with preemptive resume, absolute priority without preemptive resume, relative priority. In this situation priority systems fit into special class of countable Markov chains. Qualitative analysis of such priority queuing systems can be reduced to analysis of special classes of random walks in $\mathbf{Z}_+^\nu$. This analysis uses ideology of *induced Markov chains* and *second vector field* for the associated random walk (see [3, 5]). This approach appears to be successful in solving of many problems ([2, 3, 5]). Our goal is to find ergodicity conditions for priority systems and calculate some parameters of ergodic regime (mean waiting time, mean service time, etc.). Mathematical notion of ergodicity or stability of the queuing system corresponds to situation when mean queue lengths remain bounded when time becomes larger. This is crucial for designers of computer architecture to know that system is not overloaded: situation when queue lengths corresponding to some customer types tend to infinity means practically that newly coming customers of these types can not be served by the system in a real time.

To make our probabilistic analysis more transparent we start with simple case of two customer types, next we consider systems consisting of several infinite or finite sources and finish by compound priority systems consisting of finite and infinite sources.

Similar analysis can be done also for hierarchical model of section 2.3.

3.1 Absolute Priority Analysis in the General Setting

We consider absolute priority with preemptive resume. Assume first that all sources are infinite. We do not assume that the input flow is Poisson (but Assumption 1 is crucial). Let L_i be the mean length between subsequent arrivals of i-type demands and M_i be the mean service time. Define the load $\rho_i = \frac{M_i}{L_i}$. Let the notation be such that type i has an absolute priority with respect to type j if $i < j$. Take sufficiently large time interval $[0, T]$. During this time interval approximately $\frac{T}{L_1}$ 1-type demands will arrive and the mean service time of all these customers is $\rho_1 T$. The bus is free from 1-type customers during the time $(1 - \rho_1)T$. The same is true for i-type customers: mean number of arrived customers is $\frac{T}{L_i}$, their mean service time is $\rho_i T$. The latter fact is easy to see by induction, essential here is that the customers are served according to the preemptive resume discipline and so they "do not see" the interruption intervals. So we get the following proposition.

Proposition 1 *Assume all sources to be infinite. Then the system is ergodic*[1] *if*

$$\sum_{i=1}^{N} \rho_i < 1$$

The mean free time for all the system is

$$(1 - \sum_{i=1}^{N} \rho_i)T$$

The useful notion is the effective time for i-type customer. As the mean time when the system is accessible for i-type customers is $(1 - \sum_{j=1}^{i-1} \rho_j)T$ it can be interpreted as the time itself for i-type customer becomes $(1 - \sum_{j=1}^{i-1} \rho_j)^{-1}$ times longer. Then the following is true. We do not give the proof of this here.

Proposition 2 *The waiting time of i-type customer is*

$$W_i = (1 - \rho_1 - \dots - \rho_{i-1})^{-1} W_1^0,$$

where W_1^0 is the mean waiting time of i-type customer in case it has absolute priority (or, the only type present). Mean service time of i-type customer (taking into account possible interruptions) is

$$(1 - \rho_1 - \dots - \rho_{i-1})^{-1} M_i$$

[1] In this proposition *ergodicity* means that there exists a stationary regime in which mean lengths of queues are finite.

The case when there are also "finite" sources is more difficult but can be first analyzed in the same rough way.

3.2 Case of two customer types

We will show in this and subsequent sections that queueing systems with absolute priority disciplines under additional assumptions about input streams and service time distributions can be considered as Markov processes with state space $\mathbf{Z}_+^\nu$ (these Markov processes appears to be continuous time maximally homogeneous random walks, state $x(t)$ of the process is a vector, component $x_{ij}(t)$ of which is the number of i-type customer in the node j. In case of relative priority it is possible to consider the queueing system as a Markov process on a finite union of "octants" $\mathbf{Z}_+^\nu$. Moreover, the stochastic evolution inside an "octant" $\mathbf{Z}_+^\nu$ is also continuous time maximally homogeneous random walk. This make possible to apply new constructive methods of analysis of countable Markov chains (see [3, 5]) to studying priority queueing systems. It appears that for special class of random walks associated with priority systems it is possible to give explicit construction of the *second vector field.* This approach gives rise to constructive and practically meaningfull method of rigorous qualitative analysis of priority queuing systems.

We shall use below notations from [3, 5]. For definition of induced chain and main results concerning classification of random walks via second vector field (SVF) we refer to Chapter 4 in [3].

Our concluding remark is that analysis of continuous time random walks can be reduced to analysis of discrete time random walks (see [3, 5]). To do this one should consider imbedded discrete time random walk connected with jump-times (see section 1.3 in [3]). Qualitative behavior and stationary distribution in case of ergodicity will be the same as for original continuous time random walk. We will use below this reduction without any additional comments.

3.2.1 Infinite sources: absolute priority discipline

Let us consider the simplest example of one server node with two customer types. Customers of type 1 have priority over customers of type 2. We will suppose also that arrival streams of customers are Poisson and that the service time is exponentially distributed. Let λ_i be the input flow intensity of customers of type i, μ_i be service rate for customer of type i. Let n_1 and n_2 be the numbers of customers of types 1 and 2 in the system. So we can associate the state of the system with a point of positive quadrant $\mathbf{Z}_+^2$. Qualitative behavior of this priority system is related to the qualitative behavior of *associated* random walk on $\mathbf{Z}_+^2$ defined below.

Let us choose the following jump probabilities $p_{\alpha\beta}$.

If $\alpha \in B^{\{1,2\}} = \{(n_1, n_2) : n_1 > 0, n_2 > 0\}$ or $\alpha \in B^{\{1\}} = \{(n_1, 0) : n_1 > 0\}$ we put

$$p_{\alpha\beta} = \begin{cases} w\mu_1, & \text{if } \beta - \alpha = (-1, 0), \\ w\lambda_1, & \text{if } \beta - \alpha = (1, 0), \\ w\lambda_2, & \text{if } \beta - \alpha = (0, 1), \\ 0, & \text{for any other } \beta \neq \alpha, \end{cases} \quad , \quad p_{\alpha\alpha} = 1 - \sum_{\beta \neq \alpha} p_{\alpha\beta},$$

for all priority disciplines mentioned above.

If $\alpha \in B^{\{2\}} = \{(0, n_2) : n_2 > 0\}$ and the service discipline is *absolute priority with preemptive resume* we put

$$p_{\alpha\beta} = \begin{cases} w\lambda_1, & \text{if } \beta - \alpha = (1, 0), \\ w\lambda_2, & \text{if } \beta - \alpha = (0, 1), \\ w\mu_2, & \text{if } \beta - \alpha = (0, -1). \\ 0, & \text{for any other } \beta \neq \alpha, \end{cases} \quad , \quad p_{\alpha\alpha} = 1 - \sum_{\beta \neq \alpha} p_{\alpha\beta},$$

If $\alpha \in B^{\{2\}}$ and the service discipline is *absolute priority without preemptive resume* we put

$$p_{\alpha\beta} = \begin{cases} w\lambda_1, & \text{if } \beta - \alpha = (1, -1), \\ w\lambda_2, & \text{if } \beta - \alpha = (0, 1), \\ w\mu_2, & \text{if } \beta - \alpha = (0, -1). \\ 0, & \text{for any other } \beta \neq \alpha, \end{cases} \quad , \quad p_{\alpha\alpha} = 1 - \sum_{\beta \neq \alpha} p_{\alpha\beta},$$

For $\alpha = (0, 0)$ we put

$$p_{0\beta} = \begin{cases} w\lambda_1, & \text{if } \beta = (1, 0), \\ w\lambda_2, & \text{if } \beta = (0, 1), \\ 0, & \text{for any other } \beta \neq 0, \end{cases} \quad , \quad p_{00} = 1 - \sum_{\beta \neq 0} p_{0\beta},$$

in both cases. It is assumed that $w > 0$ is such that $w(\lambda_1 + \lambda_2 + \mu_1 + \mu_2) < 1$.

We shall give now the explicit form of *second vector field* (SVF) in both cases.

First of all we consider SVF $v^{\{1,2\}}$ for the face $B^{\{1,2\}}$:

$$v^{\{1,2\}} = w(\lambda_1 - \mu_1, \lambda_2).$$

We note that $v_2^{\{1,2\}} > 0$, so the face $B^{\{1\}} = \{(n_1, 0) : n_1 > 0\}$ is always nonergodic. Face $B^{\{2\}}$ is ergodic if $\lambda_1 < \mu_1$ and nonergodic if $\lambda_1 \geq \mu_1$. Assume that $\lambda_1 < \mu_1$. Consider the induced chain $\mathcal{L}^{\{2\}}$. This is a reversible Markov chain stationary distribution of which can be given explicitly:

$$\pi_i = \frac{1}{1 - \lambda_1/\mu_1} \left(\frac{\lambda_1}{\mu_1}\right)^i, \quad i \geq 0.$$

So in case of *absolute priority with preemptive resume* SVF for the face $B^{\{2\}}$ is equal to

$$v^{\{2\}} = w(\lambda_2 - \mu_2)\pi_0 + w\lambda_2(1 - \pi_0) = w[\lambda_2 - \mu_2(1 - \frac{\lambda_1}{\mu_1})].$$

In case of *absolute priority without preemptive resume* SVF on face $B^{\{2\}}$ is equal to

$$v^{\{2\}} = w(\lambda_2 - \mu_2 - \lambda_1)\pi_0 + w\lambda_2(1 - \pi_0) = w[\lambda_2 - (\mu_2 + \lambda_1)(1 - \frac{\lambda_1}{\mu_1})].$$

In terms of second vector field, the system of this example is ergodic if and only if $v_1^{\{1,2\}} < 0$ and $v^{\{2\}} < 0$. So priority system is ergodic iff

in the case of *absolute priority with preemptive resume*

$$\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} < 1,$$

in the case of *absolute priority without preemptive resume*

$$\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2 + \lambda_1} < 1\ .$$

3.2.2 Priority system with finite and infinite sources

Let us consider priority system with one server node (node 0 in terms of section 2.1) and customers of two types. We suppose that the system is closed with respect to the customers of type 1 and open with respect to the customers of type 2. There are N customers of type 1 and each of them can be in the buffer (at node 1 in terms of section 2.1) or waiting for the service at the server node. Each customer independently of other customers waits in the buffer exponential time with mean θ^{-1} and goes then to the server node. Service time for customers of type 1 is exponential with mean μ_1^{-1}. After service the customer of type 1 goes to the buffer. There is a Poisson input flow with intensity λ_2 of customers of type 2 to the server node, after the service customer of type 2 leaves the system. Service time of customer of type 2 is assumed to be exponential with mean value μ_2^{-1}.

We shall consider two cases:

Case 1: customers of type 1 have absolute priority with preemptive resume over customers of type 2,

Case 2: customers of type 2 have absolute priority with preemptive resume over customers of type 1.

The state space for this priority system is

$$\{(n_1, n_2): \ 0 \le n_1 \le N, \ n_2 \ge 0\} = \{0, \dots, N\} \times \mathbf{Z}_+.$$

Here n_1 is number of customers of type 1 in the queue at the server node, n_2 is number of customers of type 2 waiting for service in the queue at the server node.

As before we shall consider *associated* random walks in a half strip

$$\{0, \dots, N\} \times \mathbf{Z}_+.$$

Transition probabilities are defined as follows.

In *case 1* we put

$$p_{\alpha\beta} = \begin{cases} w(N - n_1)\theta, & \text{if } \beta - \alpha = (1, 0), \\ w\mu_1, & \text{if } \beta - \alpha = (-1, 0), \\ w\lambda_2, & \text{if } \beta - \alpha = (0, 1), \end{cases} \qquad p_{\alpha\alpha} = 1 - \sum_{\beta \ne \alpha} p_{\alpha\beta},$$

for $\alpha \in B^{\{1,2\}} \cup B^{\{1\}} = \{(n_1, n_2): \ 0 < n_1 \le N, \ n_2 \ge 0\}$ and

$$p_{\alpha\beta} = \begin{cases} wN\theta, & \text{if } \beta - \alpha = (1,0), \\ w\mu_2, & \text{if } \beta - \alpha = (0,-1), \\ w\lambda_2, & \text{if } \beta - \alpha = (0,1), \end{cases} \qquad p_{\alpha\alpha} = 1 - \sum_{\beta \ne \alpha} p_{\alpha\beta},$$

for $\alpha \in B^{\{2\}} = \{(0, n_2): \ n_2 > 0\}$.

In *case 2* we put

$$p_{\alpha\beta} = \begin{cases} w(N - n_1)\theta, & \text{if } \beta - \alpha = (1,0), \\ w\lambda_2, & \text{if } \beta - \alpha = (0,1), \\ w\mu_2, & \text{if } \beta - \alpha = (0,-1), \end{cases} \qquad p_{\alpha\alpha} = 1 - \sum_{\beta \ne \alpha} p_{\alpha\beta},$$

for $\alpha \in B^{\{1,2\}} \cup B^{\{2\}} = \{(n_1, n_2): \ 0 \le n_1 \le N, \ n_2 > 0\}$ and

$$p_{\alpha\beta} = \begin{cases} w(N - n_1)\theta, & \text{if } \beta - \alpha = (1,0), \\ w\mu_1, & \text{if } \beta - \alpha = (-1,0), \\ w\lambda_2, & \text{if } \beta - \alpha = (0,1), \end{cases} \qquad p_{\alpha\alpha} = 1 - \sum_{\beta \ne \alpha} p_{\alpha\beta},$$

for $\alpha \in B^{\{1\}} = \{(n_1, 0): \ n_1 > 0\}$.

Consider the *induced* Markov chain $\mathcal{L}^{\{2\}}$ with state space $\mathbf{Z}_N = \{0, \dots, N\}$. In *case 1* the set $\mathbf{Z}_N$ is the unique class of essential states. So according to Section 3.1 in [3] MC $\mathcal{L}^{\{2\}}$ is ergodic iff the second vector field on face $B^{\{2\}}$ is negative:

$$v^{\{2\}} = \pi_0(\lambda_2 - \mu_2)w + (1 - \pi_0)\lambda_2 w = \lambda_2 w - \pi_0 w \mu_2 < 0.$$

It is easy to check that for this case

$$\pi_0 = \frac{(\mu_1/\theta)^N}{N!} \left(1 + \frac{\mu_1}{\theta} + \frac{1}{2!}\left(\frac{\mu_1}{\theta}\right)^2 + \dots + \frac{1}{N!}\left(\frac{\mu_1}{\theta}\right)^N\right)^{-1}.$$

So in *case 1* ergodicity condition has the following form: $\dfrac{\lambda_2}{\mu_2} < \pi_0$.

In *case 2* MC $\mathcal{L}^{\{2\}}$ has only one essential state $\{N\}$. This means that priority system is ergodic iff second component $v_2^{\{1,2\}}$ of *second vector field* inside $B^{\{1,2\}}$ is negative:

$$v_2^{\{1,2\}} = w(\lambda_2 - \mu_2) < 0.$$

So in this case ergodicity condition is very simple

$$\lambda_2 < \mu_2$$

and does not depend on θ, μ_1.

3.2.3 Relative priority analysis

We shall consider here system with two customer types and one server node. Our assumptions are:

Both types 1 and 2 are infinite, input flows are Poisson, λ_1 and λ_2 are the intensities of these flows.

Type 1 has relative priority over type 2. This mean that if at the moment when customer of type 1 arrives the server is occupied by some customer of type 2 then customer of type 1 waits till customer of type 2 being served and then occupies the server node.

Service time is exponential for both types, μ_1 and μ_2 are the intensities of service.

Let us note that knowing only the number of customers of types 1 and 2 in the system is not sufficient to have Markovian description of the system. To have Markovian system we need additional information about the type of customer being served at current time moment. So the state space of our system should be chosen as follows:

$$\mathcal{S} = \{\emptyset\} \cup \{(\sigma, n_1, n_2) : \ \sigma = 1, 2, \ n_i \in \mathbf{Z}_+\}$$

where σ is a customer type being served at the server node, n_1 is a number of 1-customers waiting for service, n_2 is a number of 2-customers waiting for service; so total number of customers in the system is equal to $n_1 + n_2 + 1$. The state $\{\emptyset\}$ means that there are no customers in the system.

We are going now to construct the second vector field. Consider the induced MC $\mathcal{L}^{\{2\}}$ with state space $\mathcal{S}' = \{(\sigma, n_1) : \ \sigma = 1, 2, \ n_1 \in \mathbf{Z}_+\}$ and transition probabilities

$$p^{\{2\}}_{\gamma_1\gamma_2} = \sum_y p_{(\gamma_1,x),(\gamma_2,y)}$$

where $x > 0$.

It is convenient to consider $\mathbf{Z}^1$ as a state space of MC $\mathcal{L}^{\{2\}}$: state $k > 0$ means that there are k customers of type 1 in systems and current customer at server node has type 1; state $k \leq 0$ means that there are $-k$ customers of type 1 in systems and current customer at server node has type 2. It is easy to see that the transition probabilities are of the following form:

$$p_{k,k+1} = \lambda_1 w, \ p_{k,k-1} = \mu_1 w, \ p_{k,k} = 1 - w(\lambda_1 - \mu_1) \quad \text{for } k > 0$$

$$p_{0,-1} = \lambda_1 w, \ p_{0,0} = 1 - \lambda_1 w,$$

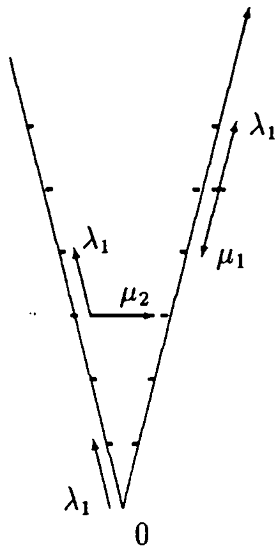

Figure 7: Induced chain $\mathcal{L}^{\{2\}}$ in case of relative priority

$$p_{k,k-1} = \lambda_1 w, \; p_{k,-k} = \mu_2 w, \; p_{k,k} = 1 - w(\lambda_1 - \mu_2) \qquad \text{for } k < 0$$

(see Figure 7).

It is easy to prove that MC $\mathcal{L}^{\{2\}}$ is ergodic iff $\lambda_1 < \mu_1$. Moreover its stationary distribution $\{\pi_k, k \in \mathbf{Z}^1\}$ can be found in explicit form by solving corresponding difference equations. It is easy to check that in case $\mu_1 \neq \lambda_1 + \mu_2$ we have the following relations

$$\pi_{-k} = \pi_0 \left(\frac{\lambda_1}{\lambda_1 + \mu_2}\right)^k,$$

$$\pi_k = \pi_0 \left[\left(\frac{\lambda_1}{\mu_1}\right)^k + \frac{\lambda_1}{\mu_1 - \lambda_2 - \mu_2}\left(\left(\frac{\lambda_1}{\lambda_1 + \mu_2}\right)^{k-1} - \left(\frac{\lambda_1}{\mu_1}\right)^{k-1}\right)\right], \quad k \geq 1.$$

Situation when $\mu_1 = \lambda_1 + \mu_2$ is the *resonance* case for the corresponding difference equation. In this case stationary probabilities satisfy the following relations

$$\pi_{-k} = \pi_0 \left(\frac{\lambda_1}{\mu_1}\right)^k, \quad \pi_k = \pi_0 k \left(\frac{\lambda_1}{\mu_1}\right)^k, \quad k \geq 1.$$

Suppose that $\lambda_1 < \mu_1$. Then the second vector field on face $B^{\{2\}}$ is equal to

$$v^{\{2\}} = w \left[\left(\sum_{k \le 0} \pi_k \right) (\lambda_2 - \mu_2) + \left(\sum_{k>0} \pi_k \right) \lambda_2 \right].$$

It is easy to calculate that in the case $\mu_1 \neq \lambda_1 + \mu_2$ we have

$$\sum_{k \le 0} \pi_k = \pi_0 \frac{\lambda_1 + \mu_2}{\mu_2},$$

$$\sum_{k>0} \pi_k = \pi_0 \left[\frac{\lambda_1}{\mu_1 - \lambda_1} + \frac{\lambda_1}{\mu_1 - \lambda_2 - \mu_2} \left(\frac{\lambda_1 + \mu_2}{\mu_2} - \frac{\lambda_1}{\mu_1 - \lambda_1} \right) \right]. \qquad (1)$$

The system under consideration is ergodic if and only if $v^{\{2\}} < 0$. So ergodicity conditions in the case $\mu_1 \neq \lambda_1 + \mu_2$ have the following form:

$$\begin{cases} \lambda_1 < \mu_1, \\ \dfrac{\lambda_1 + \mu_2}{\mu_2}(\lambda_2 - \mu_2) + \left[\dfrac{\lambda_1}{\mu_1 - \lambda_1} + \dfrac{\lambda_1}{\mu_1 - \lambda_2 - \mu_2} \left(\dfrac{\lambda_1 + \mu_2}{\mu_2} - \dfrac{\lambda_1}{\mu_1 - \lambda_1} \right) \right] \lambda_2 < 0. \end{cases}$$

If $\mu_1 = \lambda_1 + \mu_2$ then we deal with resonance case and equation (1) is not more valid. In this case we have

$$\sum_{k>0} \pi_k = \pi_0 \sum_{i=1}^{\infty} i \left(\frac{\lambda_1}{\mu_1} \right)^i = \pi_0 \frac{\lambda_1}{\mu_1} \cdot \frac{1}{\left(1 - \dfrac{\lambda_1}{\mu_1} \right)^2}.$$

So ergodicity conditions in case $\mu_1 = \lambda_1 + \mu_2$ are of the form:

$$\begin{cases} \lambda_1 < \mu_1, \\ \dfrac{\lambda_1 + \mu_2}{\mu_2}(\lambda_2 - \mu_2) + \dfrac{\lambda_1 \mu_1}{(\lambda_1 - \mu_1)^2} \lambda_2 < 0. \end{cases}$$

3.3 Several customer types of infinite type at one server node

3.3.1 Absolute priority with preemptive resume

Suppose that we have K different customer types, customers of type 1 have absolute priority (with preemptive resume) over customers of types 2,3,..., customers of type 2 have absolute priority (with preemptive resume) over customers of types 3,4,... and so on. We shall suppose also that the following assumption is true.

Assumption 4 *Parameters of the model $\lambda_i, \mu_i, i = 1, \ldots, K$, are such that for all k we have $\sum_{i=1}^{k} \frac{\lambda_i}{\mu_i} \neq 1$.*

This assumption is not very restrictive because we leave out consideration a set of parameters of zeroth Lebesgue measure.

As in case of two types, we associate with our system a random walk in $\mathbf{Z}_+^K$. We prove here that condition

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k} < 1 \tag{2}$$

is sufficient and necessary for ergodicity and give explicit description of the corresponding second vector field.

Proposition 3 (Explicit construction of the second vector field) *Second vector field enjoys the following properties:*

1. *Face $\Lambda_1 = \{1, 2, \ldots, K\}$ is outgoing for all $(K-1)$-dimensional faces $\Lambda_1 - \{i\}, i \neq 1$, and is ingoing for face $\Lambda_2 = \{2, 3, \ldots, K\}$ if $\frac{\lambda_1}{\mu_1} < 1$ and outgoing for it if $\frac{\lambda_1}{\mu_1} > 1$.*

2. *Second vector field on $(K - k + 1)$-dimensional face $\Lambda_k = \{k, k + 1, \ldots, K\}$ is equal to*

$$v^{\Lambda_k} = w\left(\lambda_k - \left(1 - \sum_{i=1}^{k-1} \frac{\lambda_i}{\mu_i}\right)\mu_k, \lambda_{k+1}, \ldots, \lambda_{K+1}\right).$$

So face Λ_k is outgoing for all $(K-k)$-dimensional faces $\Lambda_k - \{i\}, i = k+1, \ldots, K$. Face Λ_k is ingoing for face $\Lambda_{k+1} = \{k+1, \ldots, K\}$ if $\sum_{i=1}^{k} \frac{\lambda_i}{\mu_i} < 1$ and outgoing for it if $\sum_{i=1}^{k} \frac{\lambda_i}{\mu_i} > 1$.

3. *If Λ_k is ergodic denote $\sigma_{1,\ldots,k}$ the stationary probability of origin for the induced chain $\mathcal{L}^{\{1,\ldots,k\}}$. Then*

$$\sigma_{1,\ldots,k} = 1 - \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i}.$$

(In other words $\sigma_{1,\ldots,k}$ is the stationary probability of event "there are no customers of types $1, \ldots, k$ at server node" for the ergodic Markov chain obtained from whole system by removing customers of types $k+1, \ldots, K$.)

Proposition 4 *Condition (2) is sufficient and necessary for the ergodicity.*

Remark. It follows from results of subsection 3.3.2 that condition (2) is *not necessary* for ergodicity of the system with *absolute priority without preemptive resume* discipline.

Proof. It is easy to show that $\mathcal{L}^{\{2,3,\ldots,K\}}$ is ergodic if $\frac{\lambda_1}{\mu_1} < 1$ and $\sigma_1 = 1 - \frac{\lambda_1}{\mu_1}$. We are able now to calculate explicitly *the second vector field* on face $\Lambda_2 = \{2, 3, \ldots, K\}$. We get that if $\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} < 1$ then face Λ_3 is ingoing and cut-off system $\mathcal{L}^{\{1,2\}}$ obtained by removing all customers of types 3,4,..., K is ergodic. Consider the *flow balance equation* for customers of type 2:

$$\lambda_2 - (\sigma_1 - \sigma_{1,2})\,\mu_2 = 0.$$

We get immediately that

$$\sigma_{1,2} = 1 - \frac{\lambda_1}{\mu_1} - \frac{\lambda_2}{\mu_2}.$$

It is easy now to finish the proof by induction. Flow balance equation for type k has the following form

$$\lambda_k - (\sigma_{1,\ldots,k-1} - \sigma_{1,\ldots,k})\,\mu_k = 0.$$

We leave details to the reader.

Definition. The time interval between the moment when server starts service of a customer of type i and moment when this customer leaves system forever is called *real service time* for type i.

In case of absolute priority with preemptive resume real service time M_i for type i consists of time during which the customer is served at server node and of interruption intervals.

Proposition 5 *In ergodic case mean waiting time (in stationary regime) of customer of type i is equal to*

$$W_i = \left(1 - \sum_{l=1}^{i-1} \frac{\lambda_l}{\mu_l}\right)^{-1} \frac{\mu_i}{(\mu_i - \lambda_i)^2}$$

and mean real service time is equal to

$$M_i = \left(1 - \sum_{i=1}^{i-1} \frac{\lambda_i}{\mu_i}\right)^{-1} \frac{1}{\mu_i}.$$

Remark. Let W_i^0 be mean waiting time and M_i^0 be mean service time in situation when type i has highest priority. Then $W_i^0 = \mu_i/(\mu_i - \lambda_i)^2$ and $M_i^0 = 1/\mu_i$.

Let us look on behavior of ergodic system during very long time $[0, T]$. From above results and law of large numbers we get that during time of order $\left(1 - \sum_{i=1}^{i-1} \frac{\lambda_i}{\mu_i}\right) T$ system will be free of customers of types $1,2,\ldots,i-1$ and i-customers, as having highest priority, during this period of time can be served. Remaining part of time (of order $\left(\sum_{i=1}^{i-1} \frac{\lambda_i}{\mu_i}\right) T$), service of i-customers is blocked by customers with higher priority. This means that passage of i-customers through the system is $\left(1 - \sum_{i=1}^{i-1} \frac{\lambda_i}{\mu_i}\right)^{-1}$ times slower than the passing in case when i-customers have highest priority. It is natural to call factor $e_i = \left(1 - \sum_{i=1}^{i-1} \frac{\lambda_i}{\mu_i}\right)^{-1}$ *effective time* for customers of type i.

3.3.2 Absolute priority without preemptive resume

In this section we suppose that the following assumption is true.

Assumption 5 *Parameters of the model $\lambda_i, \mu_i, i = 1, \ldots, K$, are such that for all k we have* $\sum_{i=1}^{k} \frac{\lambda_i}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}} \neq 1$.

We shall prove here that in case of absolute priority without preemptive resume system of K different infinite sources is ergodic iff

$$\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2 + \lambda_1} + \frac{\lambda_3}{\mu_3 + \lambda_1 + \lambda_2} + \cdots + \frac{\lambda_K}{\mu_k + \lambda_1 + \cdots + \lambda_{K-1}} < 1. \quad (3)$$

Denote

$$S_k = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2 + \lambda_1} + \frac{\lambda_3}{\mu_3 + \lambda_1 + \lambda_2} + \cdots + \frac{\lambda_k}{\mu_k + \lambda_1 + \cdots + \lambda_{k-1}}.$$

Proposition 6 (Explicit construction of second vector field) *Second vector field enjoys the following properties:*

1. *Face $\Lambda_1 = \{1, 2, \ldots, K\}$ is outgoing for all $(K-1)$-dimensional faces $\Lambda_1 - \{i\}, i \neq 1$, and is ingoing for face $\Lambda_2 = \{2, 3, \ldots, K\}$ if $\frac{\lambda_1}{\mu_1} < 1$ and outgoing for it if $\frac{\lambda_1}{\mu_1} > 1$.*

2. *Second vector field on $(K-k+1)$-dimensional face $\Lambda_k = \{k, k+1, \ldots, K\}$ is equal to*

$$v^{\Lambda_k} = w\left(\lambda_k - (1 - S_{k-1})\mu_k, \lambda_{k+1}, \ldots, \lambda_{K+1}\right).$$

So face Λ_k is outgoing for all $(K-k)$-dimensional faces $\Lambda_k - \{i\}, i = k+1, \ldots, K$. Face Λ_k is ingoing for face $\Lambda_{k+1} = \{k+1, \ldots, K\}$ if $S_k < 1$ and outgoing for it if $S_k > 1$.

3. *If Λ_k is ergodic denote $\sigma_{1,\ldots,k}$ the stationary probability of origin for the induced chain $\mathcal{L}^{\Lambda_k}$. Then*

$$\sigma_{1,\ldots,k} = 1 - S_k \equiv 1 - \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}}.$$

Proposition 7 *Condition (3) is necessary and sufficient for ergodicity.*

Proof. Proof is similar to the proof for the previous case. The only difference consist in that flow balance condition for type k has the form

$$\lambda_k - (\mu_k + \lambda_1 + \cdots + \lambda_{k-1})(\sigma_{1,\ldots,k-1} - \sigma_{1,\ldots,k}) = 0.$$

Proposition 8 *In ergodic case mean real service time is equal to*

$$M_i = \frac{1}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}};$$

mean waiting time (in stationary regime) of customer of type i is equal to

$$W_i = (1 - S_{i-1})^{-1} \frac{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}}{(\mu_i + \lambda_1 + \cdots + \lambda_{i-1} - \lambda_i)^2}.$$

Remark. It is easy to see that in ergodic case the stationary probability of event that given customer of type i will be lost (interrupted before the end of service) is equal to

$$\frac{\lambda_1 + \cdots + \lambda_{i-1}}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}}.$$

Knowing these parameters is important for computer designer who wants to chose optimal priority order $P(i)$.

3.3.3 Fluid approximation for systems with absolute priority

Absolute priority with preemptive resume.

We are in situation of section 3.3.1 and suppose that Assumption 4 holds. Let n be such that

$$\sum_{i=1}^{n-1} \frac{\lambda_i}{\mu_i} < 1 \quad \text{but} \quad \sum_{i=1}^{n} \frac{\lambda_i}{\mu_i} > 1.$$

According to the Proposition 3 the only ergodic faces are $\Lambda_1, \ldots, \Lambda_n$. We recall that the second vector field is defined only on ergodic faces (see [3]).

Our next goal is to construct some deterministic dynamical system in $\mathbf{R}_+^K$ which is related to the original stochastic system in $\mathbf{Z}_+^K$.

We introduce the following faces in $\mathbf{R}_+^K$:

$$\tilde{\Lambda}_k = \left\{x = (x_1, \ldots, x_K) \in \mathbf{R}_+^K : \; x_1 = \cdots = x_{k-1} = 0, \; x_k > 0, \ldots, x_K > 0\right\}.$$

Consider the following vector field $v(x)$ on $\mathbf{R}_+^K$:

$$v(x) = v^{\Lambda_1} \quad \text{if} \quad x \in \tilde{\Lambda}_1 \setminus \tilde{\Lambda}_2,$$
$$\ldots$$
$$v(x) = v^{\Lambda_{n-1}} \quad \text{if} \quad x \in \tilde{\Lambda}_{n-1} \setminus \tilde{\Lambda}_n,$$
$$v(x) = v^{\Lambda_n} \quad \text{if} \quad x \in \tilde{\Lambda}_n.$$

In case than $n = K + 1$ we put $v(0) = 0$. This vector field determines dynamical system $U_t, t \geq 0$, on $\mathbf{R}_+^K$:

$$\frac{d}{dt} U_t x = v\left(U_t x\right), \quad U_0 x \equiv x.$$

This system is well defined for all $x \in \mathbf{R}_+^K$.

Let $x \in \mathbf{R}^K$. Suppose that initially we start from the configuration $[xN] \stackrel{\text{def}}{=} ([x_1 N], \ldots, [x_K N])$, i.e. we have at time $s = 0$ queue from $[x_1 N]$ 1-customers, ..., queue from $[x_K N]$ K-customers at server node. Denote by $X_s([xN])$ the state of the system at time s.

Proposition 9 (Euler limit) *For any $x \in \mathbf{R}_+^K$ the following deterministic limit exists*

$$\frac{1}{N} X_{[tN]}(xN) \to U_t x, \quad (N \to \infty) \text{ (almost sure).}$$

It is clear from the Proposition that the dynamical system U_t defined above describes evolution of large queues: if at time 0 state of priority system is $[xN]$ then at time tN state of the system will be "approximately" equal to $[(U_t x) N]$. The word "approximately" means that deviation of the true state at time tN from value $[(U_t x) N]$ is of order less than N.

There are two different cases:

($n = K + 1$) This correspond to the situation then all faces $\Lambda_1, \ldots, \Lambda_K$ and $\Lambda_{K+1} = \{0\}$ are ergodic and hence total priority system is ergodic. In this case trajectory $U_t x$ of any point $x \in \mathbf{R}_+^K$ reaches point 0 for finite time (see Figure 8a). Denote by $\tau(x)$ time of reaching origin by trajectory of dynamical system started from x:

$$U_t x \neq 0 \text{ for } t < \tau(x) \text{ and } U_t x = 0 \text{ for } t \geq \tau(x).$$

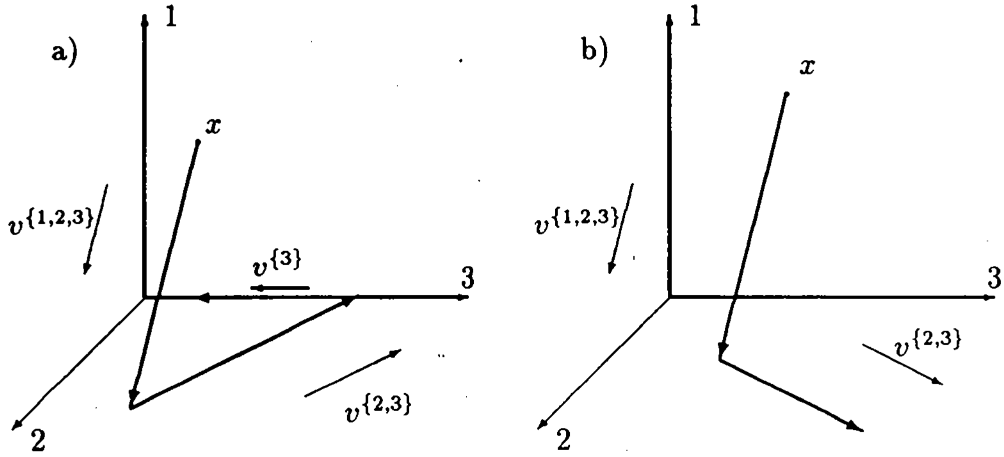

Figure 8: Dynamical system U_t in ergodic (a) and nonergodic (b) cases.

($n \leq K$) In this situation faces $\Lambda_1, \ldots, \Lambda_n$ are ergodic but all components of the second vector field on face Λ_n are strictly positive and hence total priority system is not ergodic. In this case any trajectory $U_t x$ reaches the face $\tilde{\Lambda}_n$ in finite time $\tau_n(x)$ and remains on this face forever (without going to faces of less dimension). In terms of priority systems this means that after time $\tau_n(x)N$ lengths of queues of customers of types $1, 2, \ldots, n-1$ will be of order less than N but queues of customers of types $n, n+1, \ldots, K$ will grow proportionally to N (see Figure 8b).

Now we are going to calculate the time to reach the origin $\tau(x)$ in ergodic cases.

Lemma 10 *For the ergodic system time $\tau(x)$ to reach the origin by the dynamical system U_t is linear in the initial point x. Namely,*

$$\tau(x) = (b, x) \equiv b_1 x_1 + \cdots + b_K x_K$$

where vector $b = (b_1, \ldots, b_K)$ has the following form

$$\begin{aligned}
b_1 &= e_2(1+e_3)(1+e_4)\ldots(1+e_{K+1})/\mu_1, \\
b_2 &= e_3(1+e_4)\ldots(1+e_{K+1})/\mu_2, \\
&\ldots \\
b_{K-1} &= e_K(1+e_{K+1})/\mu_{K-1}, \\
b_K &= e_{K+1}/\mu_K,
\end{aligned} \tag{4}$$

where we have denoted

$$e_k = \left(1 - \sum_{i=1}^{k-1} \frac{\lambda_i}{\mu_i}\right)^{-1}.$$

Remark. Note that quantity e_k is the same what we have called in section 3.3.1 effective time for type k.

Proof of the Lemma. Let $x \in \tilde{\Lambda}_1 \backslash \tilde{\Lambda}_2$. It is easy to check that trajectory $U_t x$ reaches face $\tilde{\Lambda}_2$ at time

$$\tau_2(x) = \frac{x_1}{\mu_1 - \lambda_1} \equiv \frac{e_2}{\mu_1} x_1.$$

Suppose that trajectory has reached face $\tilde{\Lambda}_{k-1}$ at time $\tau_{k-1}(x)$. Then it will reach face $\tilde{\Lambda}_k$ at time

$$\tau_k(x) = \frac{e_k}{\mu_{k-1}} x_{k-1} + (1 + e_k)\,\tau_{k-1}(x).$$

Note that $\tau(x) \equiv \tau_{K+1}(x)$. It is easy now to finish the proof. □

Design optimization problem Above analysis gives rise to the following *optimization problem.* We restrict ourself to ergodic case. Imagine that we are computer designers and the characteristics $\lambda_i, \mu_i, i = \overline{1,K}$ of infinite sources are known. We want to choose a priority order $i_1, i_2, \ldots, i_K$ such that the corresponding priority system starting from some "far enough" state $[xN]$ reaches equilibrium quickest possible. In other words we are looking for a permutation χ: $(1, \ldots, K) \to (i_1, \ldots, i_K)$ to ensure the smallest time $\tau_\chi(x)$ to reach origin starting from x. It is easy to see that for fixed x we should choose such priority order χ for which vector b_χ has the smallest projection on direction x.

So we come to the following definition. Let us fix some finite measure ψ on $(K-1)$-dimensional sphere $\Delta_K = \left\{x \in \mathbf{R}_+^K : x_1^2 + \cdots + x_K^2 = 1\right\}$. This measure should be chosen by the computer designer according to some external reasons which are not discussed here. If we have no reasons to give preference to some directions $\gamma \in \Delta_K$ we should take ψ equal to the uniform distribution on Δ_K .

Definition. Priority order χ is *optimal* for a given choice of measure ψ if

$$\tau_\chi(\psi) \stackrel{\text{def}}{=} \int\limits_{\Delta_K} \tau_\chi(\gamma)\, d\psi(\gamma) \;=\; \min_\sigma\; \tau_\sigma(\psi)$$

where minimum is taken over all permutations $\sigma:\ (1,\ldots,K) \to (i_1,\ldots,i_K)$.

Proposition 11 *If ψ is the uniform distribution on Δ_K then priority order χ is optimal iff vector b_χ has the smallest projection on direction* $(1,1,\ldots,1)$.

Absolute priority without preemptive resume

In the situation of section 3.3.2, suppose that Assumption 5 is true. Fluid analysis of this case is similar to the analysis for the case of absolute priority with preemptive resume. All differences are related to the fact that the second vector fields for these cases are different.

First of all we fix integer n such that

$$\sum_{i=1}^{n-1} \frac{\lambda_i}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}} < 1 \;\text{ but }\; \sum_{i=1}^{n} \frac{\lambda_i}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}} > 1.$$

It follows from Proposition 6 that the only ergodic faces are $\Lambda_1,\ldots,\Lambda_n$. On these faces second vector field is well defined. In the same way as for absolute priority with preemptive resume we define vector field $v(x)$ on $\mathbf{R}_+^K$ which generate a well defined dynamical system $U_t, t \geq 0$, on $\mathbf{R}_+^K$:

$$\frac{d}{dt}U_t x = v\left(U_t x\right), \quad U_0 x \equiv x.$$

Denote by $X_s\left([xN]\right)$ the state of the queueing system at time s provided that we start at time $s=0$ from the following configuration of queues $[xN] = \left([x_1 N],\ldots,[x_K N]\right)$.

Proposition 12 (Euler limit) *For any $x \in \mathbf{R}_+^K$ the following deterministic limit exists*

$$\frac{1}{N} X_{[tN]}(xN) \to U_t x, \quad (N \to \infty) \text{ a.s.}.$$

All subsequent analysis is similar to the case of absolute priority with preemptive resume. We want to mention only the expression for the time $\tau(x)$ to reach the origin for the trajectory $U_t x, t \geq 0$, in ergodic cases.

Lemma 13 *In ergodic case the time $\tau(x)$ to reach the origin by the dynamical system U_t is equal to*

$$\tau(x) = (b, x)$$

where the vector $b = (b_1, \ldots, b_K)$ has the form (4), where the quantities e_k are defined as follows

$$e_k = \left(1 - \sum_{i=1}^{k-1} \frac{\lambda_i}{\mu_i + \lambda_1 + \cdots + \lambda_{i-1}}\right)^{-1}.$$

3.3.4 Relative priority

As we have mentioned above this case is more delicate due to more complicated state space. State space in case of K infinite types with relative priority should be chosen as follows

$$\mathcal{S} = \{\emptyset\} \cup \{1, \ldots, K\} \times \mathbf{Z}_+^K.$$

Here the state $\{\emptyset\}$ means that there are no customer in the system, state $\{i\} \times (n_1, \ldots, n_K)$ means that there are queues of n_1 customer of type 1, ..., of n_K customers of type K and the server node is occupied by customer of type i, so total number of customers in system equals to $n_1 + \cdots + n_K + 1$. It is easy to write down the transition probabilities. This situation can be also analyzed by method of induced chains (see section 3.2.3 for introductory example).

3.4 Necessary ergodicity condition for special class of priority disciplines

We restrict ourself to the systems with infinite sources of customers of types $1, \ldots, K$. We suppose that discipline is such that the following conditions hold:

Condition 1 Service time for given customer of type k coming to the server node is determined before starting the service and is exponentially distributed with mean μ_k^{-1}.

Condition 2 No customer can leave the system before its service time is exceeded.

Remark. For absolute priority with preemptive resume and relative priority conditions 1 and 2 hold. For absolute priority without preemptive resume condition 2 is not true.

Proposition 14 (Necessary ergodicity condition) *Suppose that both conditions 1 and 2 hold and the system is ergodic. Then the following inequality holds*

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k} < 1.$$

Remark. Condition 2 in the above proposition is essential. To see this consider absolute priority discipline without preemptive resume in case

$$K = 2,\ \lambda_1 = \lambda_2 = 1,\ \mu_1 = \mu_2 = 2.$$

This system is ergodic but $\frac{\lambda_1}{\mu_1} + \frac{\lambda_1}{\mu_2} = 1$.

Proof of the proposition. Let a be the type being served at the node at current time moment. Event $\{a = \emptyset\}$ means that there are no customers in the system. Suppose that system is ergodic and P_{st} is its stationary distribution. It is evident that

$$\mathsf{P}_{st}\{a = \emptyset\} + \mathsf{P}_{st}\{a = 1\} + \cdots + \mathsf{P}_{st}\{a = K\} = 1.$$

Flow balance condition for type k has the form

$$\lambda_k - \mu_k \mathsf{P}_{st}\{a = k\} = 0.$$

So we get

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k} = 1 - \mathsf{P}_{st}\{a = \emptyset\}$$

and proposition is proved.

3.5 System of several infinite and finite sources

3.5.1 Absolute priority with preemptive resume

We consider system with customers of the following types:

$$\underbrace{1,\dots,K_1}_{I_1},\underbrace{K_1+1,\dots,K_1+L_1}_{F_1},K_1+L_1+1,\dots,\sum_{l=1}^{r}(K_l+L_l). \tag{5}$$

Customers of type i have absolute priority with preemptive resume over customers of type j if $i<j$. We assume that types

$$\underbrace{1,\dots,K_1}_{I_1},\underbrace{K_1+L_1+1,\dots,K_1+L_1+K_2}_{I_2},\dots$$

correspond to infinite sources and types

$$\underbrace{K_1+1,\dots,K_1+L_1}_{F_1},\dots,\underbrace{\sum_{l=1}^{r-1}(K_l+L_l)+K_r+1,\sum_{l=1}^{r}(K_l+L_l)}_{F_r}$$

correspond to finite sources. We assume also that intensity of input flow for infinite type $k=\sum_{i=1}^{l-1}(K_i+L_i)+K_l+v\in I_l$ is $\lambda_v^{(l)}$, service time is exponential with mean $\left(\mu_v^{(l)}\right)^{-1}$.

Consider any group F_l of finite types as separate priority system. This is an irreducible finite Markov chain which is always ergodic. Denote by $\pi_0^{(l)}$ stationary probability of event "there are no customers at server node" (all customers are in the buffer).

Suppose that the whole system is ergodic. Our goal is to get necessary ergodicity conditions. Denote $\Pi^{(l)}$ the stationary probability of event "there are no customers of types from groups $I_1,F_1,\dots,I_l$ in system". Define "total load" $R^{(l)}$ of the group I_l as

$$R^{(l)}=\sum_{i=1}^{K_l}\frac{\lambda_i^{(l)}}{\mu_i^{(l)}}.$$

The following relations hold

$$\Pi^{(l)}=\Pi^{(l-1)}\pi_0^{(l-1)}-R^{(l)},\ l\geq 2,\quad \Pi^{(1)}=1-R^{(1)}. \tag{6}$$

¿From ergodicity assumption it follows that all probabilities (6) are positive. This gives rise to the following *necessary ergodicity conditions*

$$\begin{cases} R^{(1)} < 1, \\ R^{(l)} < \Pi^{(l-1)}\pi_0^{(l-1)}, \ l = 2, 3, \ldots, r. \end{cases} \tag{7}$$

These conditions appear also to be sufficient for ergodicity.

Proposition 15 *Compound system of different finite and infinite sources with absolute priority with preemptive resume is ergodic iff conditions (7) hold.*

Our goal now is to write expressions for mean real service time and for mean waiting time for a customer of infinite type k in ergodic case. Suppose that system is in the stationary regime. For any $k \leq K_{l+1}$ consider

$$\sigma_{I^1,F^1,\ldots,I^l,F^l,1,\ldots,k} \stackrel{\text{def}}{=}$$

$$\stackrel{\text{def}}{=} \mathsf{P}\left\{ \begin{array}{c} \text{there are no customers of types from the groups } I^1, F^1, \ldots, I^l, F^l \\ \text{and of the first } k \text{ types from the group } I^{l+1} \text{ in the system} \end{array} \right\}.$$

Lemma 16

$$\sigma_{I^1,F^1,\ldots,I^l,F^l,1,\ldots,k} = \Pi^{(l)}\pi_0^{(l)} - \sum_{i=1}^{k} \frac{\lambda_i^{(l+1)}}{\mu_i^{(l+1)}}.$$

Proposition 17 *Mean waiting time for infinite type k (see notation above) is equal to*

$$W_k^{(l)} = \left(\sigma_{I^1,F^1,\ldots,I^l,F^l,1,\ldots,k}\right)^{-1} \frac{\mu_k^{(l+1)}}{\left(\mu_k^{(l+1)} - \lambda_k^{(l+1)}\right)^2}$$

and mean real service time is equal to

$$M_k^{(l)} = \left(\sigma_{I^1,F^1,\ldots,I^l,F^l,1,\ldots,k}\right)^{-1} \frac{1}{\mu_k^{(l+1)}}.$$

Analogous formulae can be written for customer of finite type.

3.5.2 Absolute priority without preemptive resume for infinite sources

Consider system of different customer types (5) with the following disciplines: if one or both of types i or j are from *finite* sources then discipline between i and j is absolute priority *with preemptive resume*; if both of types are *infinite* then discipline between i and j is absolute priority *without preemptive resume*. As usually i has priority over j if $i < j$.

In this case *conditions of ergodicity* also have form (7) where $\pi_0^{(l)}$ and $\Pi^{(l)}$ are defined in the same way as above but "loads" $R^{(l)}$ are defined as follows

$$R^{(l)} = \frac{\lambda_1^{(l)}}{\mu_1^{(l)} + \sum\limits_{i=1}^{l-1}\left(\lambda_1^{(i)} + \cdots + \lambda_{K_i}^{(i)}\right)} + \frac{\lambda_2^{(l)}}{\mu_2^{(l)} + \lambda_1^{(l)} + \sum\limits_{i=1}^{l-1}\left(\lambda_1^{(i)} + \cdots + \lambda_{K_i}^{(i)}\right)} + \cdots$$

$$\cdots + \frac{\lambda_{K_l}^{(l)}}{\mu_{K_l}^{(l)} + \lambda_1^{(l)} + \cdots + \lambda_{K_l-1}^{(l)} + \sum\limits_{i=1}^{l-1}\left(\lambda_1^{(i)} + \cdots + \lambda_{K_i}^{(i)}\right)}. \tag{8}$$

Proposition 18 *Priority system is ergodic iff conditions (7) hold where* $\Pi^{(l)}$ *are defined by (6) and* $R^{(l)}$ *are defined by (8).*

Our next goal is to write expressions for mean real service time and for mean waiting time. Suppose that system is ergodic and is in stationary regime.

Lemma 19

$$\sigma_{I^1,F^1,\ldots,I^l,F^l,1,\ldots,k} = \Pi^{(l)}\pi_0^{(l)} -$$

$$-\sum_{i=1}^{k} \frac{\lambda_i^{(l+1)}}{\mu_i^{(l+1)} + \lambda_1^{(l+1)} + \cdots + \lambda_{i-1}^{(l+1)} + \sum\limits_{i=1}^{l}\left(\lambda_1^{(i)} + \cdots + \lambda_{K_i}^{(i)}\right)}$$

Mean real service time for infinite type of order k from group I_{l+1} is equal to

$$M_k^{(l)} = \left(\mu_k^{(l+1)} + \lambda_1^{(l+1)} + \cdots + \lambda_{i-1}^{(l+1)} + \sum_{i=1}^{l}\left(\lambda_1^{(i)} + \cdots + \lambda_{K_i}^{(i)}\right)\right)^{-1}.$$

and mean waiting time is equal to

$$W_k^{(l)} = \left(\sigma_{I^1,F^1,\dots,I^l,F^l,1,\dots,k}\right)^{-1} \frac{1/M_k^{(l)}}{\left(1/M_k^{(l)} - \lambda_k^{(l+1)}\right)^2}.$$

Similar formulae can be written for customer of finite type. We leave this to the reader.

3.5.3 Second vector field and fluid approximation for compound systems

We consider the mixed model of section 3.5.1 consisting of many infinite and finite sources (5) with absolute priority with preemptive resume. We shall denote this priority system by $\mathcal{L}$. States of the system $\mathcal{L}$ are strings of the following form

$$\left(n_1, \dots, n_{K_1}, m_{K_1+1}, \dots, m_{K_1+L_1}, \dots, m_{\sum_{l=1}^{r-1}(K_l+L_l)+K_r+1}, \dots, m_{\sum_{l=1}^{r}(K_l+L_l)}\right)$$

where m_i is number of queueing customers of finite type i, n_j is number of queueing customers of infinite type j, $m_i \in \mathbf{Z}_{N_i} \equiv \{0, \dots, N_i\}$, $n_j \in \mathbf{Z}_+$. So state space of the system is

$$\mathcal{S} = \mathbf{Z}_+^{K_1} \times \mathop{\times}_{i=K_1+1}^{K_1+L_1} \mathbf{Z}_{N_i} \times \cdots \times \mathbf{Z}_+^{K_r} \times \mathop{\times}_{i=\sum_{l=1}^{r-1}(K_l+L_l)+K_r+1}^{\sum_{l=1}^{r}(K_l+L_l)} \mathbf{Z}_{N_i}.$$

Induced random walk in $\mathbf{Z}_+^{K_1+\dots+K_r}$

In this section we shall deal with *induced* Markov chain $\mathcal{L}'$ with state space

$$\mathcal{S}' = \mathbf{Z}_+^{K_1+\dots+K_r} \tag{9}$$

which we define below.

Let $\pi_0^{(l)}$ be the stationary probability of 0 for the finite MC corresponding to the group F_l of finite sources (see page 31). Let us consider first an auxiliary Markov chain $\mathcal{L}^{\text{aux}}$ with state space (9) and transition probabilities $p_{\alpha\beta}^{\text{aux}}, \alpha, \beta \in \mathcal{S}'$, which corresponds to priority system with the following infinite sources

$$\underbrace{1,\ldots,K_1}_{I_1},\underbrace{K_1+L_1+1,\ldots,K_1+L_1+K_2}_{I_2},\ldots$$

and absolute priority with preemptive resume between them. Note that if $p^{\mathrm{aux}}_{\alpha\beta} > 0$ then only one component of vector $\beta - \alpha$ can be negative (and automatically be equal to -1).

Let us define now transition probabilities $p'_{\alpha\beta}$ for induced Markov chain $\mathcal{L}'$. We put for $\alpha \neq \beta$

$$p'_{\alpha\beta} = p^{\mathrm{aux}}_{\alpha\beta} \quad \text{if} \quad \beta_j - \alpha_j \geq 0 \text{ for all } j,$$

$$p'_{\alpha\beta} = \left(\pi_0^{(1)}\pi_0^{(2)}\ldots\pi_0^{(k)}\right) p^{\mathrm{aux}}_{\alpha\beta} \quad \text{if} \quad \beta_j - \alpha_j = -1 \text{ for some } j,$$

$$\sum_{l=1}^{k}(K_l + L_l) < j \leq \sum_{l=1}^{k}(K_l + L_l) + K_{k+1},$$

and $p'_{\alpha\alpha} = 1 - \sum_{\beta\neq\alpha} p'_{\alpha\beta}$.

It appears that qualitative behavior of original Markov chain $\mathcal{L}$ is tightly connected with qualitative behavior of induced Markov chain $\mathcal{L}'$.

The following statement shows the usefulness of studying the induced chain $\mathcal{L}'$.

Proposition 20 *Markov chain $\mathcal{L}$ is ergodic if and only if Markov chain $\mathcal{L}'$ is ergodic.*

We recall that MC $\mathcal{L}$ corresponds to the *mixed* system of several *infinite* and *finite* sources but MC $\mathcal{L}'$ corresponds to the priority system of several *infinite* sources which was studied in sections 3.3.1 and 3.3.3. So the study of our original priority system $\mathcal{L}$ can be reduced to the analysis of MC $\mathcal{L}'$ which can be done in the spirit of sections 3.3.1 and 3.3.3. In particular, long time behavior of long queues of customers from infinite sources of compound system L is determined by the properties of the second vector field for the Markov chain $\mathcal{L}'$.

Second vector field

In this subsection we consider an induced MC $\mathcal{L}'$ states of which are integer valued vectors

$$(\underbrace{n_1, \ldots, n_{K_1}}_{I_1}, \underbrace{n_{K_1+1}, \ldots, n_{K_1+K_2}}_{I_2}, \ldots) \in \mathbf{Z}_+^{K_1+\cdots+K_r}.$$

Let us consider the following faces in $\mathbf{Z}_+^{K_1+\cdots+K_r}$

$$\Lambda_{l,k} = \left\{ \sum_{i=1}^{l} K_i + k, \sum_{i=1}^{l} K_i + k + 1, \ldots, \sum_{i=1}^{r} K_i \right\}.$$

Note that $\Lambda_{l,K_l+1} \equiv \Lambda_{l+1,1}$.

We give in the following proposition a full description of the second vector field for the MC $\mathcal{L}'$.

Proposition 21 (Explicit form of SVF for induced chain)

1. *The face $\Lambda_{0,1}$ is always ergodic (by definition). The second vector field on this face is equal to*

$$v^{\Lambda_{0,1}} = \left(\lambda_1^{(1)} - \mu_1^{(1)}, \lambda_2^{(1)}, \ldots, \lambda_{K_1}^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_{K_r}^{(r)}\right).$$

   So if $\lambda_1^{(1)} > \mu_1^{(1)}$ then the face $\Lambda_{0,1}$ is outgoing for any face of codimension 1; if $\lambda_1^{(1)} < \mu_1^{(1)}$ then the face $\Lambda_{0,1}$ is ingoing for $\Lambda_{0,2}$ and outgoing for any other face of codimension 1; hence the face $\Lambda_{0,2}$ is ergodic if $\lambda_1^{(1)} < \mu_1^{(1)}$.

2. *In the case when the face $\Lambda_{0,2}$ is ergodic the stationary probability of the origin for the ergodic MC $\Lambda^{\{1\}}$ is equal to $\sigma_1 = 1 - \frac{\lambda_1^{(1)}}{\mu_1^{(1)}}$.*

3. *If the face $\Lambda_{0,k}$ is ergodic then SVF on $\Lambda_{0,k}$ is given by*

$$v^{\Lambda_{0,k}} = \left(0, \ldots, \lambda_k^{(1)} - \mu_k^{(1)}\sigma_{1,\ldots,k-1}, \lambda_{k+1}^{(1)}, \ldots, \lambda_{K_r}^{(r)}\right).$$

   So if $\lambda_k^{(1)}/\mu_k^{(1)} > \sigma_{1,\ldots,k-1}$ ($\Longleftrightarrow \sum_{i=1}^{k}(\lambda_i^{(1)}/\mu_i^{(1)}) > 1$) then the face $\Lambda_{0,k}$ is outgoing for all subfaces of codimension 1; if $\sum_{i=1}^{k}(\lambda_i^{(1)}/\mu_i^{(1)}) < 1$ then the face $\Lambda_{0,k}$ is ingoing for $\Lambda_{0,k+1}$ and outgoing for any other subface of codimension 1; hence in the latter case the face $\Lambda_{0,k+1}$ is ergodic.

4. *In case when $\Lambda_{0,k}$ is ingoing for $\Lambda_{0,k+1}$ the stationary probability of the origin for the ergodic MC $\Lambda^{\{1,\dots,k\}}$ is given by*

$$\sigma_{1,\dots,k} = 1 - \sum_{i=1}^{k} \frac{\lambda_i^{(1)}}{\mu_i^{(1)}}.$$

5. *Suppose that the face $\Lambda_{l,1}$ is ergodic. Then SVF on $\Lambda_{l,1}$ is given by*

$$v^{\Lambda_{0,k}} = \left(0,\dots,\ \lambda_1^{(l+1)} - \mu_1^{(l+1)}\sigma_{I_1,\dots,I_l}\,\pi_0^{(l)},\ \lambda_2^{(l+1)},\dots,\lambda_{K_r}^{(r)}\right)$$

(here $\sigma_{I_1,\dots,I_l} \stackrel{\text{def}}{=} \sigma_{1,2,\dots,\sum_{i=1}^{l} K_i}$). So if $\lambda_1^{(l+1)}/\mu_1^{(l+1)} > \sigma_{I_1,\dots,I_l}\,\pi_0^{(l)}$ then the face $\Lambda_{l,1}$ is outgoing for all subfaces of codimension 1; if $\lambda_1^{(l+1)}/\mu_1^{(l+1)} < \sigma_{I_1,\dots,I_l}\,\pi_0^{(l)}$ then the face $\Lambda_{l,1}$ is ingoing for the face $\Lambda_{l,2}$ and is outgoing for any other subface of codimension 1; hence in the case $\lambda_1^{(l+1)}/\mu_1^{(l+1)} < \sigma_{I_1,\dots,I_l}\,\pi_0^{(l)}$ face $\Lambda_{l,2}$ is ergodic.

6. *In case when $\Lambda_{l,1}$ is ingoing for $\Lambda_{l,2}$ the stationary probability of origin for the ergodic MC $\Lambda^{\{I_1,\dots,I_l,1\}}$ is given by*

$$\sigma_{I_1,\dots,I_l,1} = \sigma_{I_1,\dots,I_l}\,\pi_0^{(l)} - \frac{\lambda_1^{(l+1)}}{\mu_1^{(l+1)}}.$$

7. *If face $\Lambda_{l,k}$ is ergodic then SVF on $\Lambda_{l,k}$ is given by*

$$v^{\Lambda_{0,k}} = \left(0,\dots,\lambda_k^{(l+1)} - \mu_k^{(l+1)}\sigma_{I_1,\dots,I_l,1,\dots,k-1},\lambda_{k+1}^{(l+1)},\dots,\lambda_{K_r}^{(r)}\right)$$

(here $\sigma_{I_1,\dots,I_l,1,\dots,k-1} \stackrel{\text{def}}{=} \sigma_{1,2,\dots,\sum_{i=1}^{l} K_i,\dots,\sum_{i=1}^{l} K_i+k-1}$). So if $\lambda_k^{(l+1)}/\mu_k^{(l+1)} > \sigma_{I_1,\dots,I_l,1,\dots,k-1}$ then the face $\Lambda_{l,k}$ is outgoing for all subfaces of codimension 1; if $\lambda_k^{(l+1)}/\mu_k^{(l+1)} < \sigma_{I_1,\dots,I_l,1,\dots,k-1}$ then face $\Lambda_{l,k}$ is ingoing for $\Lambda_{l,k+1}$ and outgoing for any other subface of codimension 1; hence in the latter case the face $\Lambda_{l,k+1}$ is ergodic.

8. *In case when $\Lambda_{l,k}$ is ingoing for $\Lambda_{l,k+1}$ stationary probability of the origin for the ergodic MC $\Lambda^{\{I_1,\dots,I_l,1,\dots,k\}}$ is given by*

$$\sigma_{I_1,\dots,I_l,1,\dots,k} = \sigma_{I_1,\dots,I_l,1,\dots,k-1} - \frac{\lambda_k^{(l+1)}}{\mu_k^{(l+1)}} \equiv \sigma_{I_1,\dots,I_l}\,\pi_0^{(l)} - \sum_{i=1}^{k} \frac{\lambda_i^{(l+1)}}{\mu_i^{(l+1)}}.$$

Fluid approximation

Consider a *deterministic* dynamical system $U_t, t \in \mathbf{R}_+$, on $\mathbf{R}_+^{K_1+\cdots+K_r}$ generated by the second vector field of MC $\mathcal{L}'$ (see section 3.3.3).

Let $Y_s = (y_1(s), y_2(s), \ldots) \in \mathcal{S}$ be the state of MC $\mathcal{L}$ at time s. Consider a random process $X_s = (y_j(s),\ j \in I_1, \ldots, I_r) \in \mathcal{S}'$ which is the "projection" of Y_s. Note that process X_s is not Markovian. Restricting ourselves to the process X_s means that we are interested only in infinite sources and do not take care about finite sources.

Let us choose the following initial configuration for $\mathcal{L}$:

$$Y_0 = (y_i^0, i \in F_1, \ldots, F_r,\ [x_j N], j \in I_1, \ldots, I_r).$$

In other words, we take *any* configuration of finite sources queues

$$y^0 = (y_i^0, i \in F_1, \ldots, F_r)$$

and configuration of long infinite sources queues

$$([x_j N], j \in I_1, \ldots, I_r),\ x \in \mathbf{R}_+^{K_1+\cdots+K_r}.$$

Denote by $X_s([xN]; y^0)$ the "projection process" X_s which corresponds to this choice of initial configuration.

Proposition 22 (Euler limit for compound system) *For any choice of y^0 the following convergence takes place for all $x \in \mathbf{R}_+^{K_1+\cdots+K_r}$*

$$\frac{1}{N} X_{[tN]}([xN]; y^0) \to U_t x,\ (N \to \infty) \quad a.s..$$

As in section 3.3.3 we are able now to calculate explicitly, the time $\tau(x)$ for the dynamical system U_t to reach the origin, starting from the point x, (this can be done, of course, only in situation when priority system is ergodic). This gives rise to the effective way for solving *optimization problem* consisting in proper choice of priority order between different customer types (see section 3.3.3).

Similar analysis can be done also for compound models of section 3.5.2.

References

[1] BASKETT F., CHANDY K.M., MUNTZ R.K., PALACIOUS F.G. *Open, closed, and mixed networks of queues with different classes of customers.* Journal of Association for Computing Machinery, Vol.22, No.2, 248-260, (1975).

[2] BOTVICH D.D., ZAMYATIN A.A. *Euler space-time limit for queuing networks.* To appear.

[3] FAYOLLE G., MALYSHEV V.A., MENSHIKOV M.V. *Topics in constructive theory of countable Markov chains.* Cambridge Univ.Press, 1994.

[4] JAISWAL, N.K. *Priority queues,* 1968.

[5] MALYSHEV V.A. *Networks and dynamical systems.* Adv.Appl.Prob., Vol.25, 140–175 (1993).

[6] VOLKOVINSKIJ M.I., KABALEVSKIJ A.N. *Analysis of priority queues.* Energoizdat, Moscow, 1981.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine - Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes - IRISA, Campus universitaire de Beaulieu 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)
Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399

RR_2419